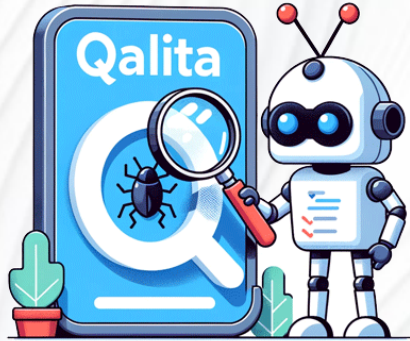


QALITA



Assurez la qualité de vos données



Une méthode claire et un accompagnement
sur mesure pour votre patrimoine data



QALITA

Platform

Fiche descriptive



QALITA Platform

Description

QALITA Platform est une solution complète permettant la mise en place, le suivi et la gestion de plans d'assurance qualité pour les données issues du Système d'Information Hospitalier (SIH). En offrant une vision claire et détaillée de la qualité des données, elle facilite leur exploitation dans les projets de recherche et pour les usages secondaires.

Bénéfices

- Gain de temps et d'efficacité dans les projets de recherche
- Mutualisation des efforts des parties prenantes sur toute la chaîne de traitement des données
- Augmentation du niveau de confiance lors de l'exploitation des données
- Amélioration continue de la gouvernance des données

Déploiement

- **On-Premise** : Contrôle total et confidentialité accrue des données
- **SaaS** : Utilisation flexible pour des données anonymisées ou en source ouverte

Gestion des Utilisateurs

Users & Roles Authentication Mode : **TABLE**

Filter login ...

User	Role	Habilitations	Created at	Last Activity
admin	Admin	No Habilitations	11 septembre 2023	11 septembre 2023
Dataengineer dataengineer	Dataengineer	No Habilitations	11 septembre 2023	13 septembre 2023
Data Scientist datascientist	Datascientist	No Habilitations	11 septembre 2023	15 septembre 2023
Datamanager datamanager	Datamanager	No Habilitations	11 septembre 2023	15 septembre 2023
public	Public	No Habilitations	11 septembre 2023	11 septembre 2023
Chief Data Officer cdo	Admin	Sensitive Data	11 septembre 2023	19 septembre 2023

Dropdown menu for 'Data Scientist' role:

- Admin
- Full Access Role
- Public
- Baseline Restricted Access Role
- Dataengineer**
- Data Engineering Access Role
- Datamanager
- Data Management Access Role
- Datascientist
- Data Scientist Access Role

Gestion facilitée des utilisateurs

- Rôles prédéfinis
- Habilitations spéciales
- LDAP / SSO / Microsoft EntraID

Dropdown menu for 'Datamanager' role:

- Sensitive Data
- Give Access To Sensitive Data
- Share Reports
- Hability To Export And Share Reports

Gestion des Utilisateurs



Connectez-vous à votre compte

Identifiant

Mot de passe

Connexion

Version 1.2.3
Copyright © - 2023 - Qalita SAS
Tous droits réservés - LICENSE

Préférences utilisateurs

- Langue EN/FR
- Home Page
- Thème de couleur

The image shows a user profile page for a 'Chief Data Officer'. The page is divided into several sections:

- Header:** A circular profile picture and the name 'Chief Data Officer'.
- Profile:** A section containing three input fields: 'Nom d'utilisateur : cdo', 'Email : null', and 'Rôle : admin'.
- Patrimoine:** A section showing a list of items categorized by role. Under 'Auteur', there are 'Projets' (1), 'Rapports' (0), and 'Tickets' (14). Under 'Propriétaire', there are 'Sources' (0) and 'Packs' (0).
- Paramètres:** A section for user preferences, including:
 - 'Langue : Français' (with a dropdown arrow)
 - 'Page d'accueil : Data Management - Projects' (with a dropdown arrow)
 - 'Thème :' with two options: 'Clair' (light theme) and 'Sombre' (dark theme).
- Jetons d'API:** A section stating 'Votre jeton API est généré et expire dans un an. Ne partagez ce jeton avec personne.'

Gestion des Partenaires

Organization Qalita You have 5 publicly available packs.

Partners

Filter name ... + Add A Partner

Partner	Created At	Last Synchronization	
Dev Qalita Platform	9 octobre 2023	16 octobre 2023	
Demo Qalita	9 octobre 2023	16 octobre 2023	

Ajoutez des partenaires pour engager la collaboration

- **Partage** de packs
- Projets **Multicentriques**

Add a Partner ×

Share your token

Share your token with partners to let them share pack with your organization. And let you read their public pack :

Or paste token from a partner

Paste the token Given to you by your partner in the input below :

Register Partner

CLOSE































Gestion des Sources de données

- Ajoutez **tout type** de sources
- Gérez leur **visibilité** (publiques, internes, privées)
- Tag spécifique pour les sources « **sensibles** »
- Gestion des sources de **référence**
- **Couverture qualité** en un coup d'œil

sources /

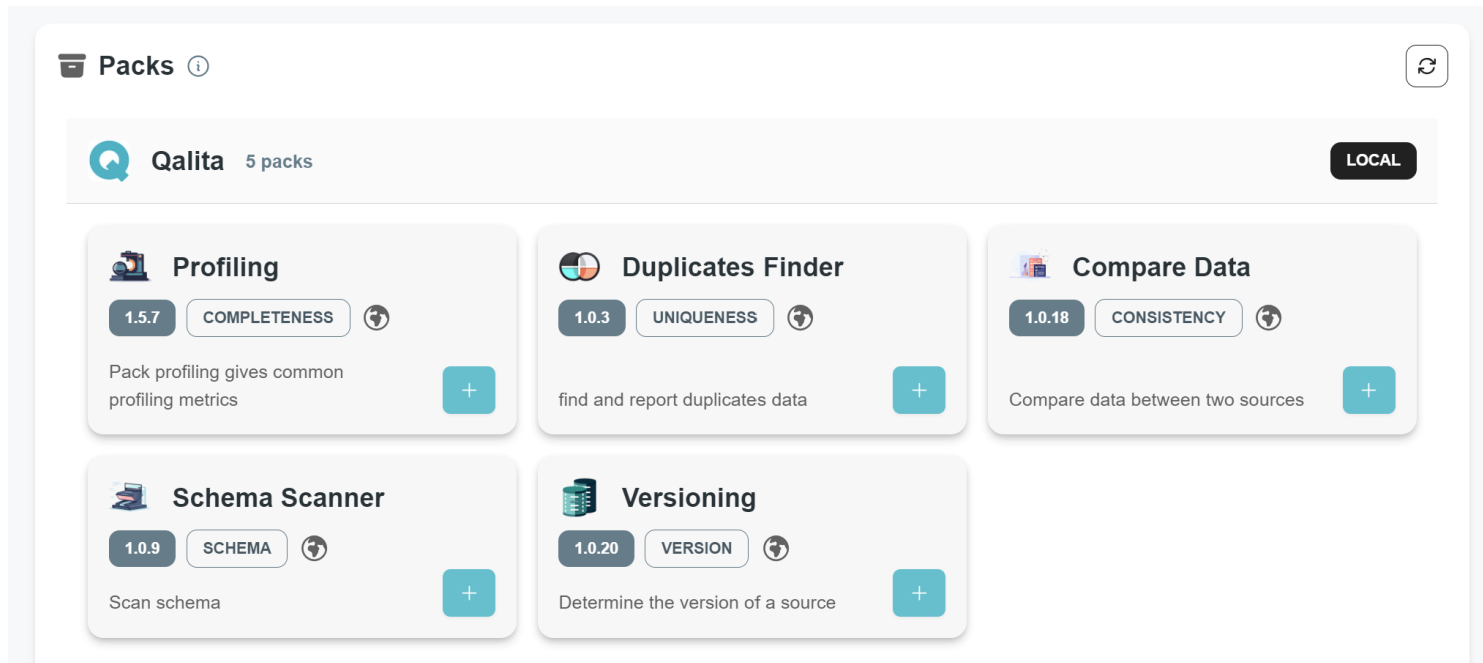
Sources 1 Sources publiées sur la plateforme : 10/30 ↻

Filter name ... + Type + Visibilité + De référence + Sensible

Nom	Couverture Qualité	Recommandations
 Early Stage Diabetes Risk Prediction Dataset <i>Sign and Symptom Data of Newly Diabetic or would be Diabetic Patient</i>		5 85 
 Breast Cancer Gene Expression Profiles (METABRIC) <i>Clinical attributes, m-RNA levels z-score, and genes mutations for 1904 patients</i>		
 Hepatitis C Prediction Dataset <i>Laboratory values of blood donors and Hepatitis C patients</i>		3 27 12 
 Heart Disease And Stroke Prevention <i>Provides a comprehensive image for cardiovascular diseases & related prevention</i>		18 174 
 Sequence And Meta Data For Various Protein Structures <i>Structural Protein Sequences</i>		28 
 Latest Covid-19 India Statewise Data <i>Covid-19 India Statewise Data as on January 17, 2023</i>		18 6 5 
 Medical Cost Personal Datasets <i>Insurance Forecast by using Linear Regression</i>		1 4 20 
 Subway Nutrition <i>All items on Subway's Menu w/ Nutrition, Micronutrients, and Calorie Information</i>		11 
 Life Longevity Factors <i>What factors will really increase your average life expectancy and lifespan?</i>		20 16 
 Heart Failure Prediction Dataset <i>11 clinical features for predicting heart disease events.</i>		7 1 60 

Les Packs

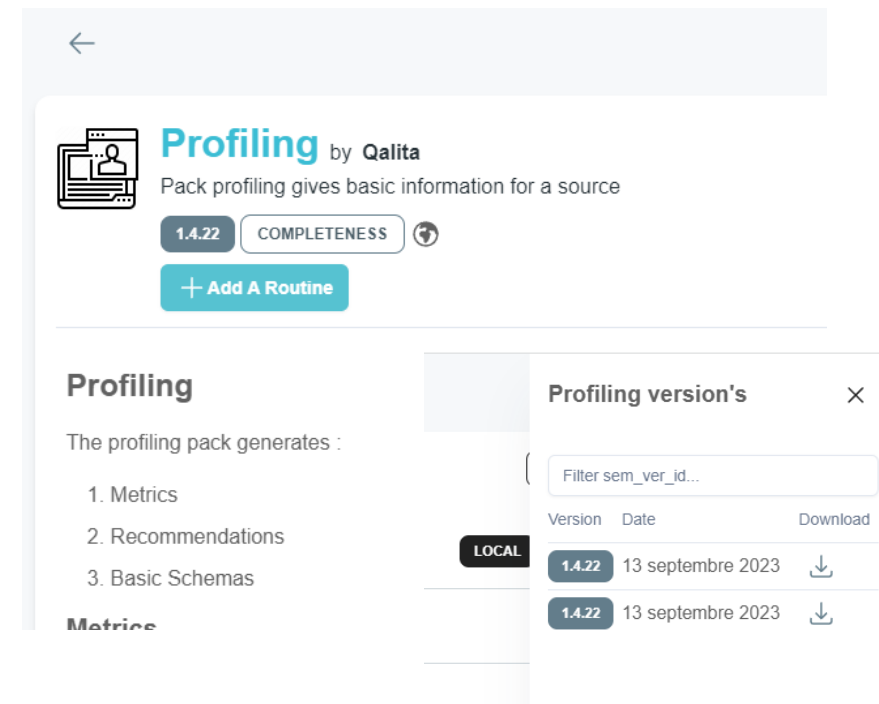
- **Adaptez** vos analyses sur mesure avec le principe de microprogrammes spécifiques aux sources de données et aux dimensions de la qualité (pack)
- **Partagez-les** avec vos partenaires pour gagner en efficacité lors de projets multicentriques



Packs ⓘ

Qalita 5 packs LOCAL

- Profiling** (1.5.7) - COMPLETENESS
 - Pack profiling gives common profiling metrics
- Duplicates Finder** (1.0.3) - UNIQUENESS
 - find and report duplicates data
- Compare Data** (1.0.18) - CONSISTENCY
 - Compare data between two sources
- Schema Scanner** (1.0.9) - SCHEMA
 - Scan schema
- Versioning** (1.0.20) - VERSION
 - Determine the version of a source



Profiling by Qalita

Pack profiling gives basic information for a source

1.4.22 COMPLETENESS

+ Add A Routine

Profiling

The profiling pack generates :

1. Metrics
2. Recommendations
3. Basic Schemas

Metrics

Profiling version's		
Version	Date	Download
1.4.22	13 septembre 2023	↓
1.4.22	13 septembre 2023	↓

Les Agents et routines

- **Orchestrez** vos analyses en ajoutant des agents en 3 étapes simples
- **Planifiez** votre production d'analyses grâce aux routines

Agent Configuration

Terminal
Docker

Step 1 : Python

Make sure you have installed a 3.8+ Python version and that you have installed the Qalita Agent package.

```
pip install qalita
```

Step 2 : Env Variables

Get your agent configuration (name, URL, mode and token) in your clipboard to copy in your terminal, this will create an .env file containing your credentials. You will be able to get your token at anytime in your user profile.

```
agentName=cdo
fileName="$HOME/.qalita/.env-$agentName"
mkdir -p $(dirname $fileName)
```

Step 3 : Source

Copy the lines below in your Terminal to get you started.

```
qalita agent -n cdo login
```

More info or documentation for the qalita agent can be found at [...](#)

Agents

Admin 37 jobs and 37 routines ● OFFLINE Since: about 8 hours ago

Jobs

⊕ Pack
⊕ Source
⊕ Status
Succeeded
Reset X

Pack	Source	Start ↑	Duration	Status
Duplicates Finder	Heart Disease And Stroke Prevention	2023-10-16	0h 0m 5s	SUCCEEDED
Compare Data	Heart Disease And Stroke Prevention	2023-10-16	0h 0m 6s	SUCCEEDED
Duplicates Finder	Latest Covid-19 India Statewise Data	2023-10-16	0h 0m 5s	SUCCEEDED
Compare Data	Latest Covid-19 India Statewise Data	2023-10-16	0h 0m 4s	SUCCEEDED
Profiling	Latest Covid-19 India Statewise Data	2023-10-16	0h 0m 26s	SUCCEEDED
Profiling	Subway Nutrition	2023-10-16	0h 0m 29s	SUCCEEDED
Duplicates Finder	Heart Failure Prediction Dataset	2023-10-16	0h 0m 5s	SUCCEEDED
Compare Data	Heart Failure Prediction Dataset	2023-10-16	0h 0m 5s	SUCCEEDED
Profiling	Heart Failure Prediction Dataset	2023-10-16	0h 0m 24s	SUCCEEDED
Duplicates Finder	Subway Nutrition	2023-10-16	0h 0m 5s	SUCCEEDED

Rows per page: 10 Page 1 of 4

Routines

Versioning → Heart Disease And Stroke Prevention

■ ● 🔧 🗑️

Duplicates Finder → Heart Disease And Stroke Prevention

■ ● 🔧 🗑️

Compare Data → Heart Disease And Stroke Prevention

■ ● 🔧 🗑️

New Routine

Pack

Profiling

Source

Select Source

Schedule

CRON Expression

30 0 * * 1,6

Every week on MON,SAT

at 0 : 30 Clear

Start date

19/09/2023

Configuration

You may need to check the pack details to see what configuration is required and know more about it.

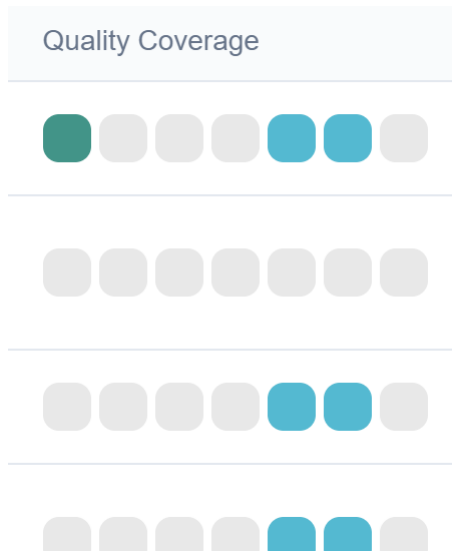
```
{
  "metrics": true,
  "metrics_chart_mapping": [
    {
      "metric_key": "missing_cells",
      "chart_type": "line_chart"
    },
    {
      "metric_key": "number_of_observations",
      "chart_type": "line_chart"
    }
  ]
}
```

Create

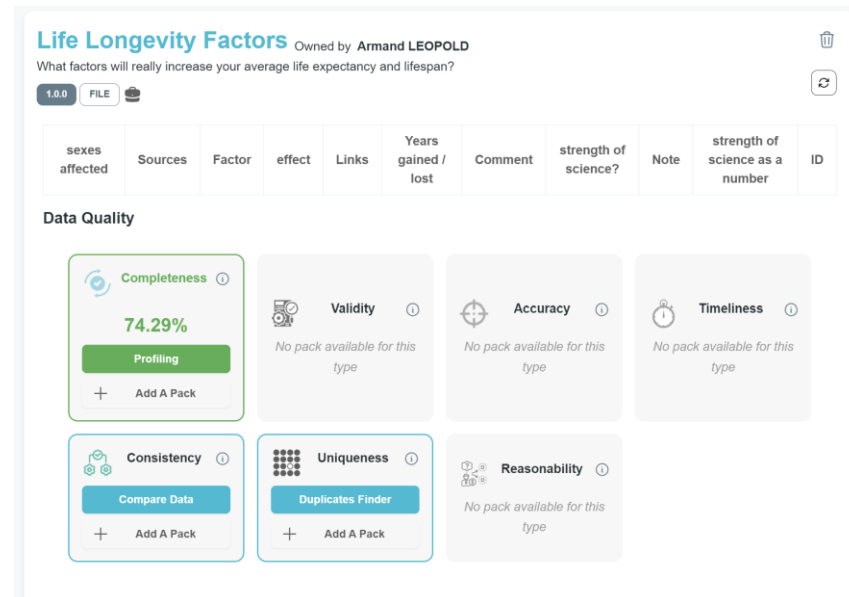
La Qualité

Plusieurs niveaux de détails pour visualiser la qualité de la donnée

- Quality Coverage Progress Bar
- Source Quality Dimension Dashboard
- Per dimension Quality metrics



Quality Coverage progressbar



Source quality dimension dashboard



Per dimension quality metrics dashboard

Les Tickets

Convertissez les recommandations en **tickets**, et prenez en main votre plan d'assurance qualité

- Suivi sur les **projets**
- Vue **Kanban** pour la gestion courante

- Plan d'actions **simple et rapide**

Heart Failure Prediction Dataset

11 clinical features for predicting heart disease events.

1.0.0 FILE

Age	Cholesterol	Oldpeak	Sex
	Zeros INFO Cholesterol has 172 (18.7%) zeros	Zeros INFO Oldpeak has 368 (40.1%) zeros	

N Cells Missing

- n_cells_missing

New issue [X]

Source: Heart Failure Prediction Dataset

Due date: 16/10/2023

Description: Cholesterol has 172 (18.7%) zeros

Create

Tickets [TABLEAU DE TICKETS]

OPEN

- Lot Of Duplicates**: Column 'CHOL' has a duplication rate of 48.94%. Consider reviewing for data cleaning. 13 avril 2024
- Unique**: Factor has unique values. 26 novembre 2023
- Lot Of Duplicates**: Column 'Discharge Ratio' has a duplication rate of 13.89%. Consider reviewing for data cleaning. 19 janvier 2024
- Info**: ALP has 18 (2.9%) missing values. 22 mars 2024

TODO

- Info**: ID has unique values. 30 novembre 2023
- Info**: ALP has 18 (2.9%) missing values. 19 avril 2024

DOING

- Info**: children has 574 (42.9%) zeros. > Les données sont à compléter. 30 novembre 2023
- Lot Of Duplicates**: Column 'CREA' has a duplication rate of 80.98%. Consider reviewing for data cleaning. 30 novembre 2023
- Lot Of Duplicates**: Column 'class' has a duplication rate of 99.62%. Consider reviewing for data cleaning. 23 novembre 2023

REVIEW

- Info**: Oldpeak has 368 (40.1%) zeros. 26 janvier 2024
- Lot Of Duplicates**: Column 'Alopecia' has a duplication rate of 99.62%. Consider reviewing for data cleaning. 23 novembre 2023

DONE

- Zeros**: Active Ratio has 32 (88.9%) zeros !! important !! 1 novembre 2023
- Lot Of Duplicates**: Column 'Age' has a duplication rate of 90.19%. Consider reviewing for data cleaning. 11 novembre 2023

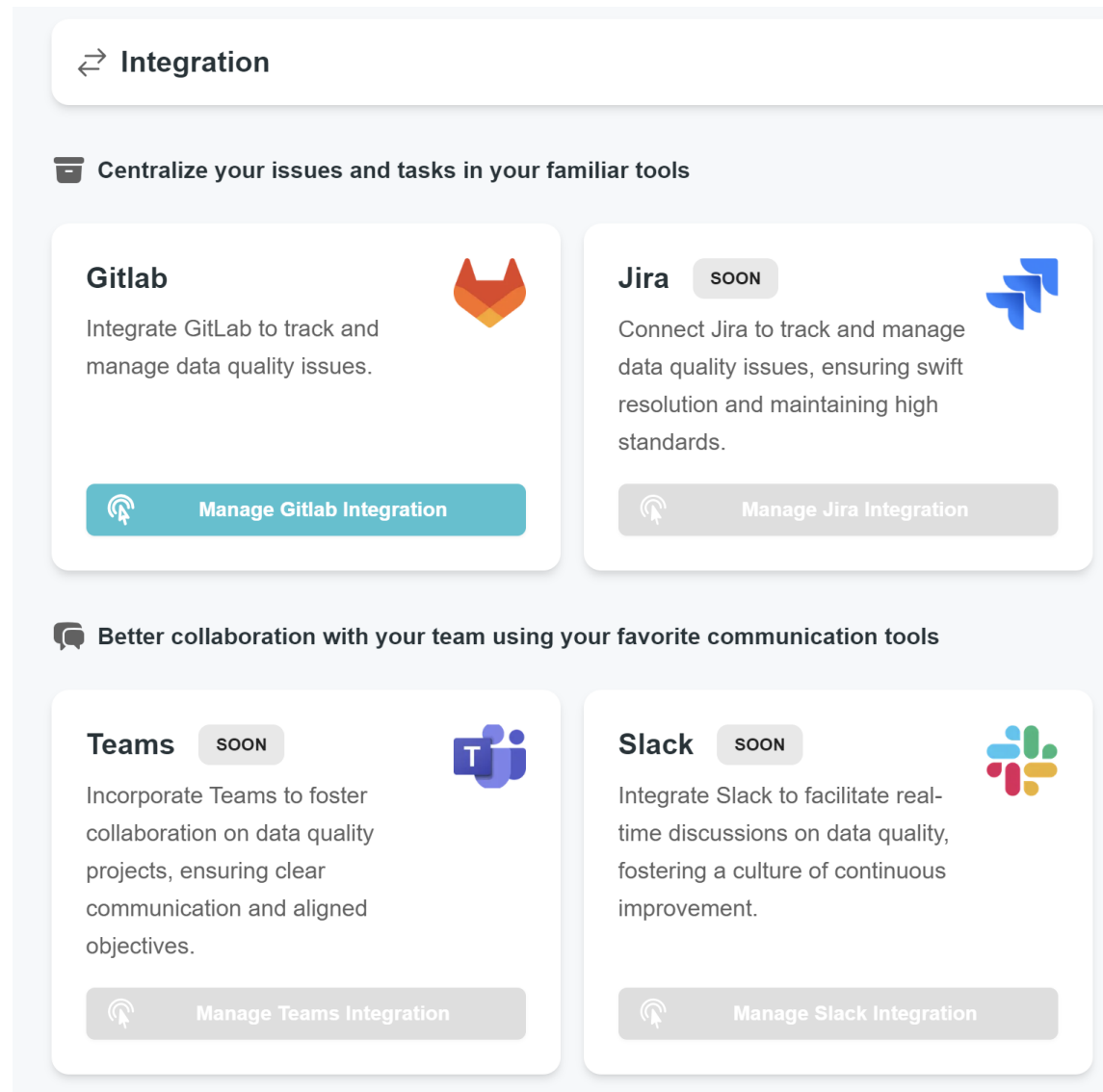
CLOSED

- Missing**: Comment has 9 (25.7%) missing values
- Lot Of Duplicates**: Column 'Active Ratio' has a duplication rate of 88.89%. Consider reviewing for data cleaning.
- Lot Of Duplicates**: Column 'BreakoutCategoryID' has a duplication rate of 100.0%. Consider reviewing for data cleaning.

Intégration

Boostez votre productivité en interconnectant Qalita avec vos outils

- **Gitlab/Jira** : Gérez vos tickets directement là ou vous en avez l'habitude
- **Teams/Slack** : Communiquez efficacement sur les sujets qualité




The screenshot shows the 'Integration' section of the QALITA interface. It features a header with a double-headed arrow icon and the word 'Integration'. Below this, there are two main sections: 'Centralize your issues and tasks in your familiar tools' and 'Better collaboration with your team using your favorite communication tools'. Each section contains two integration cards. The first card is for Gitlab, with a 'Manage Gitlab Integration' button. The second card is for Jira, with a 'Manage Jira Integration' button and a 'SOON' label. The third card is for Teams, with a 'Manage Teams Integration' button and a 'SOON' label. The fourth card is for Slack, with a 'Manage Slack Integration' button and a 'SOON' label. Each card includes a description of the integration and the respective tool's logo.

↔ Integration

🗑 Centralize your issues and tasks in your familiar tools

Gitlab

Integrate GitLab to track and manage data quality issues.




[Manage Gitlab Integration](#)

Jira

SOON

Connect Jira to track and manage data quality issues, ensuring swift resolution and maintaining high standards.




[Manage Jira Integration](#)

💬 Better collaboration with your team using your favorite communication tools

Teams

SOON

Incorporate Teams to foster collaboration on data quality projects, ensuring clear communication and aligned objectives.




[Manage Teams Integration](#)

Slack

SOON

Integrate Slack to facilitate real-time discussions on data quality, fostering a culture of continuous improvement.



[Manage Slack Integration](#)

Projets

Uniformisez votre approche par des processus clairs

🏠 / projects / Diabète

←

Di

Diabète

Géré par **Datamanager**
projet de recherche sur le diabète

🗑️

🗄 Sources

Early Stage Diabetes Risk Prediction Dataset

1.0.0
FILE
🗑️

Medical Cost Personal Datasets

1.0.0
FILE
🗑️

📄 Rapports

report 3

2023-11-21

🔗
⋮

Dataengineer

🎫 Tickets

+ Statut
+ Source
+ Assigné

	Titre ↑↓	Description	Date d'échéance	Source	Assigné
DONE	Lot Of Duplicates	Column 'Age' has a duplication rate of 90.19%. Consider reviewing for data cleaning.	11 novembre 2023	Early Stage Diabetes Risk Prediction Dataset	🔗 🗨️ 🗑️
DOING	Lot Of Duplicates	Column 'class' has a duplication rate of 99.62%. Consider reviewing for data cleaning.	23 novembre 2023	Early Stage Diabetes Risk Prediction Dataset	🔗 🗨️ 🗑️
REVIEW	Lot Of Duplicates	Column 'Alopecia' has a duplication rate of 99.62%. Consider reviewing for data cleaning.	23 novembre 2023	Early Stage Diabetes Risk Prediction Dataset	🔗 🗨️ 🗑️
DONE	Info	children has 574 (42.9%) zeros. > Les données sont à	30 novembre	Medical Cost Personal	🔗 🗨️ 🗑️

Rapports

Partagez efficacement le niveau de qualité de vos données grâce aux rapports d'analyses

Rapports

Filter title ... Projet

Titre	Projet	Créé à	Dernière mise à jour
Rapport Initial <i>Premier rapport pour partage du 21 11 2023</i>	Diabète <i>projet de recherche sur le diabète</i>	21 novembre 2023	23 novembre 2023

Partager le rapport

Vous pouvez partager votre rapport avec d'autres utilisateurs en leur donnant ce lien de partage.

Public

<https://demo.platform.qalita.io/share/3/?token=...>

- Sachez que le partage public de rapports peut exposer des données sensibles, soyez prudent lorsque vous partagez des rapports publiquement.
- Vous pouvez partager des rapports avec des utilisateurs connectés avec un rôle public, cela garantit que seuls les membres de votre organisation peuvent accéder au rapport. Pour cela, sélectionnez [Privé]

Qalita Demo

Rapport Initial

Premier rapport pour partage du 21 11 2023

Auteur : Datamanager

Créé à : 21 novembre 2023

Dernière mise à jour : 23 novembre 2023

Early Stage Diabetes Risk Prediction Dataset

Duplicates Finder

class	Obesity	Alopecia	muscle stiffness	partial paresis	delayed healing	Irr
Lot of duplicates HIGH <small>Column 'class' has a duplication rate of 99.82%. Consider reviewing for data cleaning.</small>	Lot of duplicates HIGH <small>Column 'Obesity' has a duplication rate of 99.82%. Consider reviewing for data cleaning.</small>	Lot of duplicates HIGH <small>Column 'Alopecia' has a duplication rate of 99.82%. Consider reviewing for data cleaning.</small>	Lot of duplicates HIGH <small>Column 'muscle stiffness' has a duplication rate of 99.82%. Consider reviewing for data cleaning.</small>	Lot of duplicates HIGH <small>Column 'partial paresis' has a duplication rate of 99.82%. Consider reviewing for data cleaning.</small>	Lot of duplicates HIGH <small>Column 'delayed healing' has a duplication rate of 99.82%. Consider reviewing for data cleaning.</small>	Lot of duplicates HIGH <small>Column 'Irr' has a duplication rate of 99.82%. Consider reviewing for data cleaning.</small>

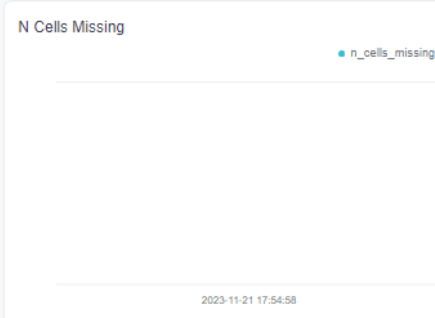
Medical Cost Personal Datasets

Profiling

charges	region	smoker	bmi	sex	age	children
						info <small>children has 574 (42.9%) zeros</small> INFO

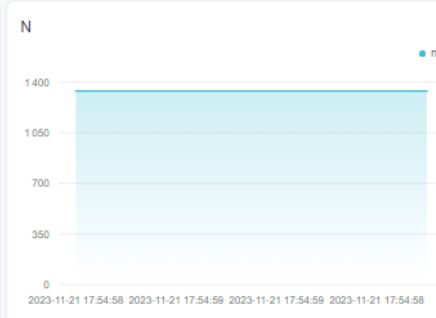
Métriques

N Cells Missing



N Cells Missing: **0**

N



N: **1400**

N Cells Missing

0

P Cells Missing

0.0

Duplicates Finder

Take action !

Test our demo instance

demo.platform.qalita.io



Pourquoi choisir QALITA ?

- **Expertise** reconnue en gestion de la qualité des données
- Solutions **adaptées** aux besoins spécifiques des acteurs de la santé
- Approche **pédagogique** et stratégique pour une mise en œuvre efficace

Contactez-nous dès aujourd'hui pour en savoir plus sur nos solutions !

contact@qalita.io





Fiche descriptive



QALITA Audit

Description

Le service QALITA Audit propose une analyse approfondie de la qualité des données d'un jeu de données ou d'un périmètre défini chez le client. L'audit identifie les écarts par rapport aux normes et bonnes pratiques, détecte les opportunités d'optimisation et fournit un rapport détaillé pour améliorer la gouvernance et la fiabilité des données.

Bénéfices

- Évaluation complète de la qualité des données (cohérence, complétude, duplication, conformité)
- Diagnostic précis avec indicateurs clés
- Recommandations stratégiques pour améliorer la gestion des données
- Présentation pédagogique pour faciliter la prise de décision

Livrables

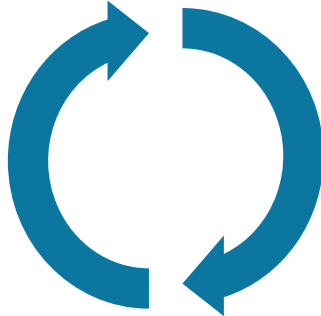
- Rapport détaillé des analyses
- Recommandations adaptées aux besoins du client
- Présentation des résultats aux équipes

Méthodologie



Déploiement

Installation de l'outil selon deux modalités (On Premise, ou SaaS).



Définition périmètre

Référencement des sources de données dans l'application.



Profilage

Utilisation de Packs spécifiques.



Analyse qualité

Production des rapports selon les dimensions prédéfinies.



Identification anomalies

Identification d'anomalies qui vont donner lieu à des tickets.



Actions correctives

Plan d'actions à partir des tickets.

Pourquoi choisir QALITA ?

- **Expertise** reconnue en gestion de la qualité des données
- Solutions **adaptées** aux besoins spécifiques des acteurs de la santé
- Approche **pédagogique** et stratégique pour une mise en œuvre efficace

Contactez-nous dès aujourd'hui pour en savoir plus sur nos solutions !

contact@qalita.io





Fiche descriptive



QALITA Formation

Description

QALITA Formation est une prestation de formation en qualité des données visant à renforcer les compétences des utilisateurs sur les bonnes pratiques et outils de gestion de la qualité. Grâce à une approche personnalisée, elle combine théorie, ateliers pratiques et évaluation des compétences.

Bénéfices

- Amélioration de la fiabilité des données et conformité réglementaire
- Optimisation des processus internes
- Augmentation de la productivité et réduction des coûts liés aux erreurs
- Meilleure intégration de la qualité des données dans les workflows existants

Formats Disponibles

- **E-Learning** : Formation en ligne flexible et accessible
- **On-Site** : Sessions sur site adaptées aux besoins spécifiques des organisations.

Variété de supports

The screenshot shows the QALITA Documentation website. The header includes the QALITA logo and 'Documentation'. A navigation menu on the left lists sections like Introduction, Démarrage, Guides, Administration, Ingénierie des données, etc. The main content area is titled 'Guides' and contains a sub-section 'Cas d'usages de la plateforme par utilisateur type'. Below this, there are five cards representing different user types: Administration (5 éléments), Ingénierie des données (4 éléments), Gestion des données (4 éléments), Gouvernance des données (2 éléments), and Science des données (2 éléments). Navigation arrows are visible at the bottom of the content area.

The screenshot shows a GitHub repository for 'qalita-io / tutorials'. The repository is public and has 1 branch and 0 tags. The README file is open, showing the title 'QALITA Tutorials' and the QALITA logo (Data Quality Platform). The introduction text reads: 'Welcome to the Qalita tutorials! Here you will find a collection of tutorials to help you get started with Qalita.' There is also a 'Table of Contents' section. The right sidebar shows repository statistics: 1 star, 0 watching, 0 forks, and a report repository link.

Vidéos ▶ Tout lire

The video thumbnails are as follows:

- Exploiter l'analyse de son patrimoine data** (2:56) - QALITA Platform : Exploiter l'analyse de son patrimoine... (6 vues • il y a 1 mois)
- Gérer les droits des utilisateurs, analyser les évènements** (5:44) - QALITA Platform : Gérer les droits des utilisateurs,...
- Utiliser le catalogue** (2:46) - QALITA Platform : Utiliser le catalogue (9 vues • il y a 2 mois)
- Exp d'ana** - QALITA F les résultat (16 vues • i)

<https://doc.platform.qalita.io/>
<https://www.youtube.com/@qalita-io>
<https://github.com/qalita-io/tutorials>

Pourquoi choisir QALITA ?

- **Expertise** reconnue en gestion de la qualité des données
- Solutions **adaptées** aux besoins spécifiques des acteurs de la santé
- Approche **pédagogique** et stratégique pour une mise en œuvre efficace

Contactez-nous dès aujourd'hui pour en savoir plus sur nos solutions !

contact@qalita.io

